



LAWRENCE  
LIVERMORE  
NATIONAL  
LABORATORY

LLNL-TR-514271

# Modeling Temporal Behavior in Large Networks: A Dynamic Mixed-Membership Model

R. Rossi, B. Gallagher, J. Neville, K. Henderson

November 16, 2011

## **Disclaimer**

---

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

# Modeling *Temporal Behavior* in Large Networks: A Dynamic Mixed-Membership Model

Ryan Rossi  
Jennifer Neville

Purdue University  
{rossi, neville}@purdue.edu

Brian Gallagher  
Keith Henderson

Lawrence Livermore Lab  
{bgallagher, keith}@llnl.gov

## ABSTRACT

Given a large time-evolving network, how can we model and characterize the temporal behaviors of individual nodes (and network states)? How can we model the behavioral transition patterns of nodes? We propose a temporal behavior model that captures the “roles” of nodes in the graph and how they evolve over time. The proposed *dynamic behavioral mixed-membership model* (DBMM) is scalable, fully automatic (no user-defined parameters), non-parametric/data-driven (no specific functional form or parameterization), interpretable (identifies explainable patterns), and flexible (applicable to dynamic and streaming networks). Moreover, the interpretable behavioral roles are generalizable, computationally efficient, and natively supports attributes. We applied our model for (a) identifying patterns and trends of nodes and network states based on the temporal behavior, (b) predicting future structural changes, and (c) detecting unusual temporal behavior transitions. We use eight large real-world datasets from different time-evolving settings (dynamic and streaming). In particular, we model the evolving mixed-memberships and the corresponding behavioral transitions of Twitter, Facebook, IP-Traces, Email (University), Internet AS, Enron, Reality, and IMDB. The experiments demonstrate the scalability, flexibility, and effectiveness of our model for identifying interesting patterns, detecting unusual structural transitions, and predicting the future structural changes of the network and individual nodes.

## Categories and Subject Descriptors

H.2.8 [Database Applications]: Data Mining

## General Terms

Algorithms, Experimentation

## Keywords

Graph mining, dynamic network models, non-negative matrix factorization, anomaly detection.

## 1. INTRODUCTION

In recent years, we have witnessed a tremendous growth in both the variety and scope of network datasets. In particular, network datasets often record the interactions and/or transactions among a set of entities—for example, personal communication (e.g., email, phone), online social network interactions (e.g., Twitter, Facebook), web traffic between servers and hosts, and router traffic among autonomous systems. A notable characteristic of these *activity* networks, is that the structure of the networks change over time (e.g., as people communicate with different friends). These temporal dynamics are key to understanding system *behavior*, thus it is critical to model and predict the network changes over time. An improved understanding of temporal patterns will facilitate for example, the development of software systems to optimally manage data flow, to detect fraud and intrusions, and to allocate resources for growth over time.

Although some recent research has focused on the analysis of dynamic networks [18, 4, 6, 13, 5, 22], there has been less work on developing *models* of temporal behavior in large scale network datasets. There has been some work on modeling temporal events in large scale networks [3, 31] and other work that uses temporal link patterns to improve predictive models [25, 24]. In addition, there is work on identifying clusters in dynamic data [6, 26] but these methods focus on discovering underlying communities over time—sets of nodes that are highly clustered together. In contrast, we are interested in uncovering the *behavioral* patterns of nodes in the network and modeling how those patterns change over time. The recent work on dynamic mixed-membership stochastic block models (dMMSB: [10, 30]), is to our knowledge, one of the only methods suitable for modeling node-centric behavior over time. The dMMSB model identifies groups of nodes with similar patterns of linkage and characterizes how group memberships change over time. However, dMMSB assumes a specific parametric form where the groups are defined through linkage to specific nodes (i.e., in particular types of groups) rather than considering more general forms of node behavior over dynamic node sets. More importantly the dMMSB estimation algorithm is not scalable, which makes the method unsuitable for analysis of large graphs.

In this work, we aim to develop a descriptive model to answer the following questions for dynamic network datasets:

1. **Identify dynamic patterns in node behavior.** What types of high level temporal patterns and trends do the data exhibit? Are behaviors cyclical or predictable? Do nodes have different behavioral patterns?

2. **Predict future structural changes.** Can we predict when a node’s role will change (e.g., a node with high in-degree transitions to a node with high betweenness)? Is the overall structure of the graph becoming more or less predictable over time?
3. **Detect unusual transitions in behavior.** Are there nodes and/or timesteps with significantly different behavioral patterns? What types of structural properties are most anomalous? Are there unusual time periods of activity or are the anomalies consistent over time?

To facilitate the investigation of these questions, we propose to model node “roles” and how they change over time. Specifically, to focus on node behavior (rather than community finding) we use non-parametric feature-based roles from the recent work of [16]. Using these non-parametric roles, which can incorporate both attributes and link structure, and generalize to new unseen nodes, we propose a novel *dynamic behavioral mixed-membership model* (DBMM) suitable for large temporal networks. The DBMM discovers features (i.e., using the graph and intrinsic attributes), extracts these features for all timesteps, and automatically learns behavioral “roles” for nodes at each timestep. The number of behavioral roles are selected automatically using MDL or AIC. Afterwards, we learn a node’s individual role transition probabilities (i.e., given a node role  $r_i$ , what is the probability of transitioning to  $r_j$  in the next timestep).

Our proposed DBMM technique allows us to investigate the properties of temporal networks and understand both global and local behaviors, detect anomalies, as well as predict future structural changes. The main strengths of the approach include:

- ★ **Automatic.** The algorithm doesn’t require user-defined parameters.
- ★ **Scalable.** The learning algorithm is linear in the number of edges in the time-interval under consideration. It is also easily parallelizable as features, roles, transition models can be discovered/extracted from each timestep independently.
- ★ **Non-parametric and data-driven.** The model structure (i.e., number of parameters) and more generally the parameterization depends on the properties of the time-evolving network.
- ★ **Interpretable and intuitive.** The DBMM is based on an intuitive behavioral representation (structural properties) of the network and individual nodes. It identifies explainable patterns, trends, and aids in understanding the underlying dynamic process.
- ★ **Flexible.** The algorithm is applicable for all types of time-evolving networks (e.g., dynamic and streaming networks).

We demonstrate the application of our model on several real world datasets, showing that it both accurately predicts future structural changes as well as identifying interesting temporal patterns and anomalies. We discuss the scalability of the approach and notably we apply the DBMM to networks with up to 300,000 nodes and 4 million edges—datasets that are orders of magnitude larger than could be modeled with dMMSBs.

## 2. DYNAMIC BEHAVIORAL MIXED MEMBERSHIP MODEL (DBMM)

Our goal is to model the behavioral roles of nodes and their evolution over time. Given a sequence of network snapshots (graphs and attributes), the Dynamic Behavioral Mixed Membership Model (DBMM) consists of (1) automatically learning a set of representative features, (2) extracting features from each graph, (3) discovering behavioral roles (4) iteratively extracting these roles from the sequence of network snapshots over time and (5) learning a predictive model of how these behaviors change over time.

### 2.1 Data Model for Temporal Networks

Given a dynamic network  $D = \{N, E\}$ , where  $N$  is the set of nodes and  $E$  is the set of edges in  $D$ , a network snapshot  $S_t = \{N_t, E_t\}$  is a subgraph of  $D$  where  $E_t$  are the edges in  $E$  active at time  $t$  and  $N_t$  are the endpoints of the edges  $E_t$ .

### 2.2 Representing Network Behavior

The idea is to discover a set of underlying roles, which together describe the behaviors observed in the network, and then assign a probability distribution over these roles to each node in the network, which explain that node’s observed behavior. Roles are extracted via a two-step process.

**Feature Discovery.** The first step is to represent and describe each active node in a given timestep using a carefully selected set of features, which capture both (a) the local structural characteristics of the node (e.g., degree, clustering) and (b) the characteristics of neighboring nodes. For this task, we leverage a recently proposed method [15] to generate a set of representative features automatically. We discover a set of features at time  $t$  denoted  $V_t$  such that  $V_t$  is an  $|N_t| \times F$  matrix where  $|N_t|$  is the number of active nodes and  $F$  is the number of features. The features are extracted for each network snapshot resulting in a sequence of node-feature matrices, denoted  $\mathbf{V} = \{V_t : t \in T\}$ .

**Role Discovery.** The next step is to automatically discover groups of nodes (representing common patterns of behavior) based on their features. For this purpose, we use Non-negative Matrix Factorization (NMF) to extract roles [16] and extend it for a sequence of graphs. Given a sequence of node-feature matrices, we generate a rank- $r$  approximation  $G_t F \approx V_t$  where each row of  $G_t \in \mathbb{R}^{n \times r}$  represents a node’s membership in each role and each column of  $F \in \mathbb{R}^{r \times f}$  represents how membership of a specific role contributes to estimated feature values. For constructing the “closest” rank- $r$  approximation we use NMF because of interpretability and efficiency, though any other method for constructing such an approximation may be used instead (SVD, spectral decomposition). Given a nonnegative matrix  $V_t \in \mathbb{R}^{n \times f}$  and a positive integer  $r < \{n, f\}$ , find nonnegative matrices  $G_t \in \mathbb{R}^{n \times r}$  and  $F \in \mathbb{R}^{r \times f}$  that minimize,

$$f(G_t, F) = \frac{1}{2} \|V_t - G_t F\|_F^2$$

The number of roles is automatically selected using Minimum Description Length (MDL) criterion. We iteratively estimate the node-role memberships for each network snapshot  $\mathbf{G} = \{G_t : t \in T\}$  given  $F$  and  $\mathbf{V} = \{V_t : t \in T\}$  using NMF. Afterwards, we have a sequence of node-role matrices  $\mathbf{G}$  where each active node at time  $t$  is represented with their current role memberships.



## 2.3 Behavioral Transition Model

Given a sequence of dynamic behaviors  $\mathbf{G} = \{G_t : t \in T\}$ , we can learn a model of how behavior in our network changes over time. More formally, given two behavioral snapshots,  $G_{t-1}$  and  $G_t$ , we can learn a transition matrix  $T_{r \times r}$  that approximates the change in behavior from time  $t-1$  to  $t$ . The transition matrix  $T$  represents how likely a node is to transition from role  $r_i$  to role  $r_j$  for that particular time interval:

$$T = \begin{bmatrix} z^{(r_1 \rightarrow r_1)} & z^{(r_1 \rightarrow r_2)} & \dots & z^{(r_1 \rightarrow r_m)} \\ z^{(r_2 \rightarrow r_1)} & z^{(r_2 \rightarrow r_2)} & \dots & z^{(r_2 \rightarrow r_m)} \\ \vdots & \dots & \ddots & \dots \\ z^{(r_m \rightarrow r_1)} & z^{(r_m \rightarrow r_2)} & \dots & z^{(r_m \rightarrow r_m)} \end{bmatrix}$$

where  $T$  is estimated using NMF such that  $G_{t-1}T \approx G_t$ .

In the simple form of the model presented above, we learn  $T$  using only a single transition (i.e.,  $t-1$  to  $t$ ). However, we can conceive of variations that leverage more available data by considering multiple transitions (*stacked modes*) or that smooth over a sequence of transitions using kernel functions (*summary model*). We discuss these variants in detail next.

### 2.3.1 Stacked Transition Model

The stacked model uses the training examples from the  $k$  previous timesteps. More formally, the stacked model is defined as,

$$\begin{bmatrix} G_{t-1} \\ G_{t-2} \\ \vdots \\ G_{k-1} \end{bmatrix} T \approx \begin{bmatrix} G_t \\ G_{t-1} \\ \vdots \\ G_k \end{bmatrix}$$

where  $k = t - w$  and  $w$  is the window size; typically  $w = 10$ . We denote the stacked behavioral snapshots as  $G_{k:t}$  where  $k:t$  represents all the training examples from timestep  $k$  to timestep  $t$ .

### 2.3.2 Summary Transition Model

This class of models uses  $k$  previous timesteps to weight the training examples at time  $t$  using some kernel function. The exponential decay and linear kernels are used in this work. The temporal weights can be viewed as probabilities that a node behavior is still active at the current time step  $t$ , given that it was observed at time  $(t - k)$ . We define the summary behavioral snapshot  $G_{S(t)}$  as a weighted sum of the temporal role-memberships up to time  $t$  as follows,

$$G_{S(t)} = \alpha_1 G_k + \dots + \alpha_{w-1} G_{t-1} + \alpha_w G_t = \sum_{i=k}^t K(G_i; t, \theta)$$

where the  $\alpha$  weights determine the contribution of each temporal snapshot in the summary model.

In addition to exponential and linear kernels, we experimented with the inverse linear and also tried various  $\theta$  values. Overall, we found the linear kernel (and exponential) to be the most accurate with  $\theta = 0.7$ . Nevertheless, the optimal  $\theta$  will depend on the type of dynamic network and the volatility.

### 2.3.3 Discussion & Observations

We have found the summary model to be the best performer for prediction tasks because of its ability to smooth over multiple timesteps. However, for precisely this reason,

the summary model is more difficult to interpret. Therefore, we use the summary model for prediction tasks and the stacked representation for data analysis tasks, due to its interpretability.

In addition, we experimented with other variants of the DBMM transition model, including a stacked-summary hybrid and multi-state models, which make an explicit distinction between transitions from activate states and transitions from inactive states. However, we opted in favor of the simpler stacked and summary models because none of these other models provided an obvious advantage.

## 3. DATASETS & ANALYSIS

We validate our dynamic behavioral mixed-membership model on a variety of temporal networks from different domains. The datasets are briefly described below and further details are provided in Table 1.

**Twitter Relationships.** We generate a social “who-follows-whom” network by starting with a few seed users (all famous people with verified accounts) and crawling “follows” links (extracted in 2008). Then from this social network, we extract tweets observed during the year. To construct a graph from tweets, we generate “mentions” edges. That is, if a tweet by user1 contains a username (e.g. @user2), we add a link between user1 and user2 in the network, but only if user2 is the first user mentioned in the tweet. We do not include self-mentions.

**Twitter (Copenhagen).** The network is formed by the set of 74,227 reply-to-messages in the #cop15 Twitter hashtag occurring over a two-week period from 12/07/09-12/18/09.

**Email (University).** The data consists of University emails corresponding to two weeks. We generated the email network using email logs from the university mailservers. We only consider email accounts that have at least one incoming or outgoing edge in the trace.

**Enterprise Network Traces.** We use real network-trace data sets collected over time on an enterprise network. The nodes are IP addresses and links are communications between IPs. Each communication has a begin-time and an end-time.

**Facebook.** This dataset consists of wall communications between users that belong to a city [28].

**Enron.** This dataset contains the email communications in Enron Inc. from Jan 1999 to July 2002 [1]. We construct sender-receiver graphs on a weekly basis.

**Internet.** To study the most recent evolutionary characteristics of the Internet, we obtained time series of BGP routing tables from the Oregon RouteViews project [21]. We extracted AS-level subgraphs for the years 2004 to 2011.

**IMDB.** The Internet Movie Database consists of movies, people, and studios connected by links (from 1980 to 2007). A link is created between actors if they acted together.

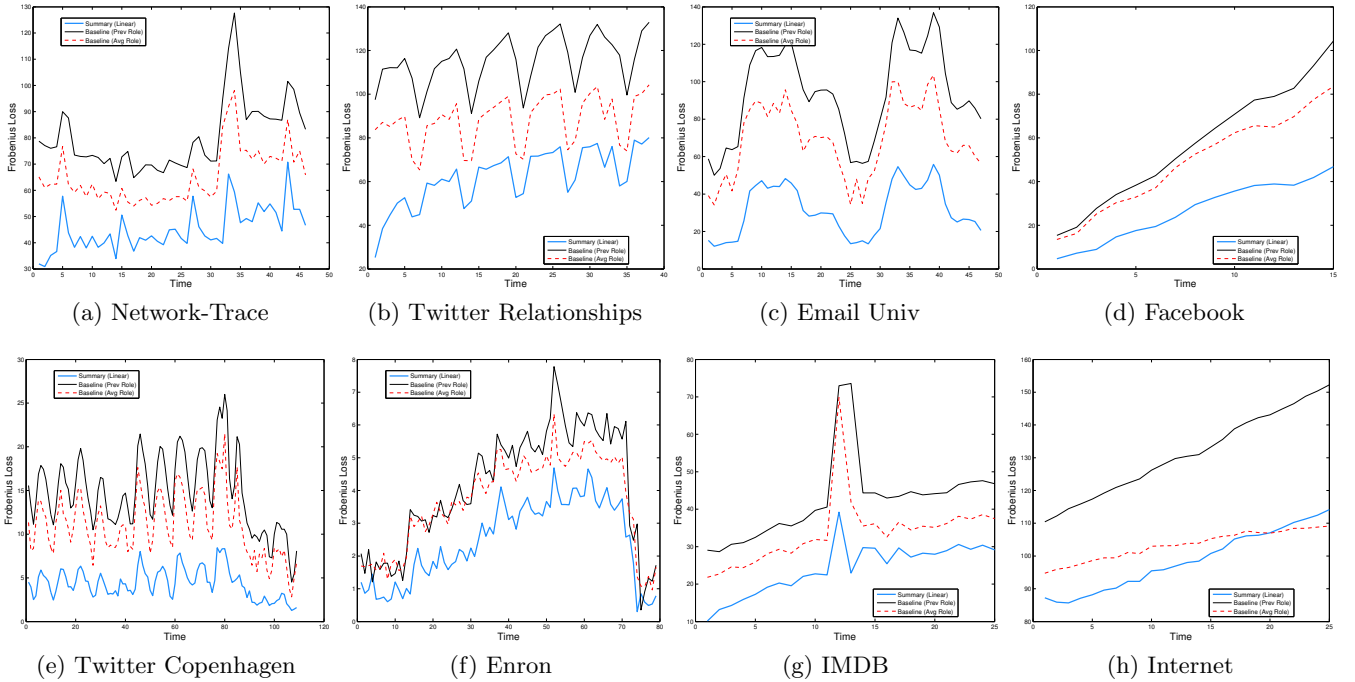
**Reality Mining.** This dataset consists of mobile-phone communications of 94 human subjects at the MIT Media Laboratory from July 2004 to June 2005.

## 3.1 Behavioral Analysis

How can we quantify the amount of randomness or complexity present in a graph? How can we quantify this randomness in terms of the underlying process that generated

**Table 1: The number of learned features, discovered behavioral roles, time intervals, and the total number of nodes is listed for each dataset. The number of features and roles provide intuition about the underlying generative process and also indicates the amount of randomness (or complexity) present in the network.**

Dataset	#Features	#Roles	N	E	T	length
TWITTER RELATIONSHIPS	1325	12	310,809	4,095,627	41	1 day
TWITTER (COPENHAGEN)	150	5	8,581	27,889	112	3 hours
FACEBOOK	161	9	46,952	183,831	18	1 day
EMAIL-UNIV	652	10	116,893	1,270,285	50	60 min
NETWORK-TRACE	268	11	183,389	1,631,824	49	15 min
INTERNET AS	30	2	37,632	505,772	28	3 months
ENRON	173	6	151	50,572	82	2 weeks
IMDB	45	3	21,257	296,188	28	1 year
REALITY	99	5	97	31,694	46	1 month



**Figure 1: The DBMM transition model accurately predicts future behavior of individual nodes (i.e., mixed role membership) compared to sensible baseline models.**

it? We analyze simple measures generated from our behavioral representation such as the number of learned features and the number of behavioral roles. Table 1 shows the number of features and behavioral roles for each dataset (among other useful information). A key observation is that the more complex or random the network structure, the more features and behavioral roles discovered. If the network structure is highly predictable or only consists of certain patterns (e.g., possibly a hierarchical structure), then these structures (or motifs) can be represented with a smaller amount of features. For instance, the Internet AS topology has some hierarchical structure or recurring patterns of connectivity among ISPs and therefore only 30 features are discovered while searching the space of recursive features (correlated/redundant features are pruned). This is in contrast to networks with more complex patterns of connectivity, i.e., twitter and transactional networks such as the email network. In these cases, the links are instantaneous

and might only last for some duration of time, making more complex structures more likely.

## 4. EXPERIMENTAL EVALUATION

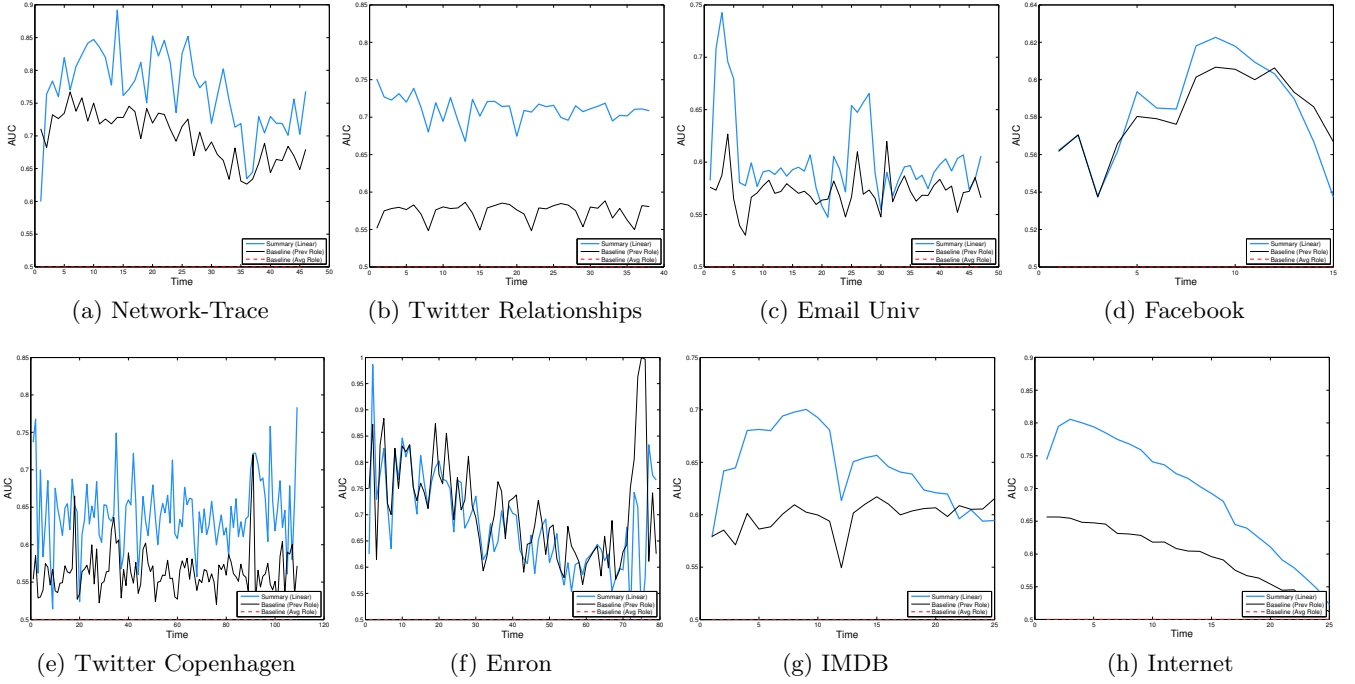
In this section, we show the utility of the proposed behavioral model for time-evolving network analysis tasks (exploratory analysis, predicting temporal behavior and identifying unusual behaviors).

### 4.1 Predicting Temporal Behavior

In this section, we validate the DBMM model by demonstrating its ability to predict the future behavior of nodes.

#### 4.1.1 Models

The goal is to accurately predict  $G_{t+1}$  given  $G_{s(t)}$ , the summary behavioral snapshot described in Section 2.3.2. Our primary means of predicting  $G_{t+1}$  is using our DBMM summary transition model  $T$  as follows:  $\hat{G}_{t+1} = G_t T$ . We



**Figure 2: The DBMM transition model accurately predicts future behavior of individual nodes (i.e., modal role membership) compared to sensible baseline models.**

compare this summary model to two sensible baselines: *PrevRole* and *AvgRole*. *PrevRole* simply assigns each node the role distribution from the previous timestep  $t$ . That is,  $\hat{G}_{t+1} = G_t$ . *AvgRole* assigns each node the average role distribution over all nodes at time  $t$ . The *AvgRole* model can be expressed as  $\hat{G}_{t+1} = G_t T_A$  where  $T_A$  is estimated from  $G_t = [1] T$ . Essentially, *PrevRole* assumes node behavior does not change from timestep to timestep and *AvgRole* assumes that all nodes exhibit the average behavior of the network.

#### 4.1.2 Evaluation

We consider two strategies for evaluating our predictive models: (a) compare the predicted  $\hat{G}_{t+1}$  to the true  $G_{t+1}$  using a loss function (we use the Frobenius norm) and (b) Use  $\hat{G}_{t+1}$  to predict the modal role of each node at time  $t+1$  and evaluate these predictions using a multi-class AUC (Area Under the ROC curve) measure. We describe each of these strategies more formally below.

**Frobenius Loss on  $G$ .** The goal here is to estimate  $G_{t+1}$  as accurately as possible. The approximation error between the estimated node memberships  $\hat{G}_{t+1} = G_t T_{t+1}$  and the true node memberships  $G_{t+1}$  is defined as

$$\|G_{t+1} - \hat{G}_{t+1}\|_F$$

**Structural Prediction with Multi-class AUC.** This is a multi-class classification task where the true class label for node  $i$  is the modal role from the  $i$ th row of  $G_{t+1}$  (i.e., the role with maximum membership for this node). The predicted class label for node  $i$  is the modal role from the  $i$ th row of  $\hat{G}_{t+1}$ .

We evaluate the predictions using a generalization of AUC extended for multi-class problems. In particular, we compute the AUC of all combinations of labels and take the

mean (also known as Total AUC) [14]. The difficulty of the prediction task varies based on the number of roles discovered, complexity of the network evolution, and the type of time-evolving network (e.g., transactional vs. social network).

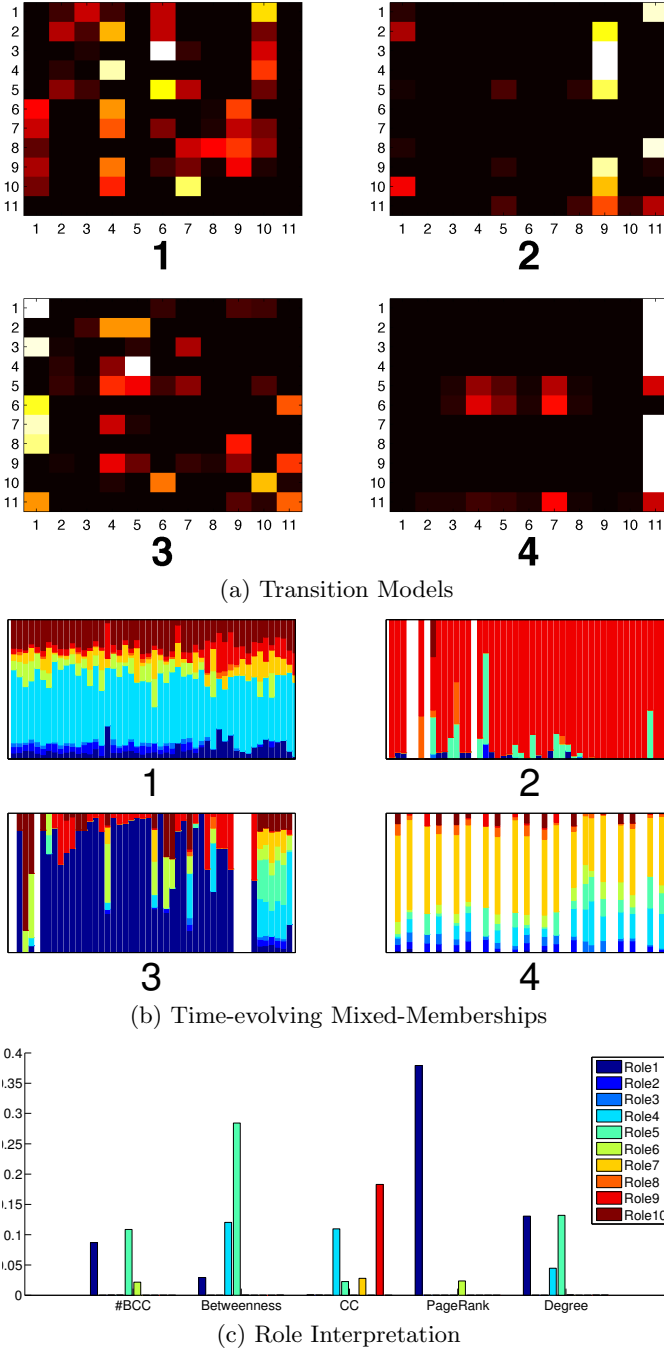
#### 4.1.3 Results

Figures 1 and 2 both demonstrate that the DBMM summary transition model is an effective predictor across the range of experiments. With few exceptions, DBMM outperforms both baselines for all data sets and timesteps. This is even true for the more complex time-evolving networks such as Twitter, email, and the IP-traces, which are more transactional with rapidly evolving network structure. Some results are omitted for space, but all are qualitatively similar.

In addition to validating the DBMM model, both figures offer some interesting insights into the characteristics of time-evolving networks. For example, an increase in loss over time may indicate a “concept drift” where behavior in the network has evolved to the point where the current roles can no longer adequately explain node behavior. This effect is seen most prominently in Figures 1(b), 1(d), 1(g) and 1(h). Interestingly, the drift we see in Figure 1(h) agrees with the current understanding that the underlying evolutionary process of the Internet AS is not constant, as was previously believed [20, 29]. Most notably, there is recent evidence of the Internet topology transitioning from hierarchal to a flat topological structure [8, 7].

Furthermore, the figures provide insights into behavioral anomalies, such as the spike we see in Figure 1(g). The spike in loss indicates the significant deviation of the node roles at a specific time.

Finally, in the large Twitter Relationships network, we see seasonality among the role transitions. In particular, we



**Figure 3: The DBMM transition model effectively captures the diverse temporal behavior of hosts in a computer network. (a) Transition matrices for 4 hosts (b) Corresponding role-memberships over time (c) Characteristics of individual roles. The y-axis of 3(a) represents the role we transition from, the x-axis the role we transition to. The x-axis of 3(b) represents time while the y-axis represents the proportion of that behavioral role at each timestep. Each distinct color represent a behavioral role.**

find that the users generally behave significantly different over the weekends, seen by the increase in loss and the decrease in AUC on these days. Intuitively, we would expect

users to be tweeting about different topics and using the system in different manner than they do during the work days. The Twitter Copenhagen network captures the more locally-temporal seasonality; that is users behave differently during the daytime and the nighttime hours.

## 4.2 Exploratory Analysis

We apply the temporal behavioral model to understand individual node network behavior and unexpected conditions (i.e., anomalies). In particular, what are the characteristics of nodes with respect to their estimated behavioral roles? Does a node’s behavior change slowly over time (days, weeks) or do nodes change behaviors cyclically (e.g., diurnal). Can we cluster nodes based on their patterns of dynamic structural behavior? In this section, we show how DBMM can be used for interpreting the behavior of nodes, roles, and their evolution (time-varying information) and identify interesting patterns and trends.

### 4.2.1 Interpretation & Analysis

We start with an illustrative example of applying DBMM to a large IP trace network, shown in Figure 3. We first plot the time-evolving mixed-memberships from four nodes shown in Figure 3(b) and then visualize their corresponding transition models in Figure 3(a). In the time-evolving mixed-memberships, inactivity is represented by white bars whereas in the transition models inactivity corresponds to the last row/column. The transition models are learned using the stacked representation which aids in the understandability and interpretation of the roles and their modeled transitions.

The time-evolving mixed-memberships for each of the four example nodes in Figure 3(b) show distinct patterns from one another which are easy to identify. The four patterns represented by these nodes can be classified as having the following patterns of structural behavior,

1. **Structural Stability.** This node’s structural behavior (and communication pattern) is relatively stable over time.
2. **Homogeneous.** The node for the most part takes on a single behavioral role.
3. **Abrupt transition.** Their structural behavior changes abruptly. In the IP-Communication network, it could be that the IP was released and someone was assigned it or perhaps that the machine was compromised and began acting maliciously.
4. **Periodic activity.** The node has periodic activity, but maintains similar structural behavior. In the case of the IP-communication network, this machine could be infected and every 30 minutes sends out a communication to the master indicating that it is connected and “listening”.

For the four example nodes, we show their transition models in Figure 3(a). The transition models represent the probability of transitioning or taking on the structural behavior of role  $j$  given that your current role (or main role) is role  $i$ . For instance, node 2 homogeneously takes on the red role over time as discussed previously. From Figure 3(c), we see that the red role is “role 9”, and looking back at the node’s learned transition model, we find that column 9 contains most of the mass, which represents that there is a high probability of transitioning from any other role to the red role. As shown in the mixed-memberships over time, this is exactly

what is expected. As another example, we find that node 4 usually transitions from a mix of active roles to the inactive role (i.e., the inactive role is represented by column/row eleven). Therefore, we would expect our learned transition model to capture this by placing most of the mass on the last column, representing the probability of going inactive after having a mix of active roles in the previous timestep, which is exactly what we see in the fourth transition model.

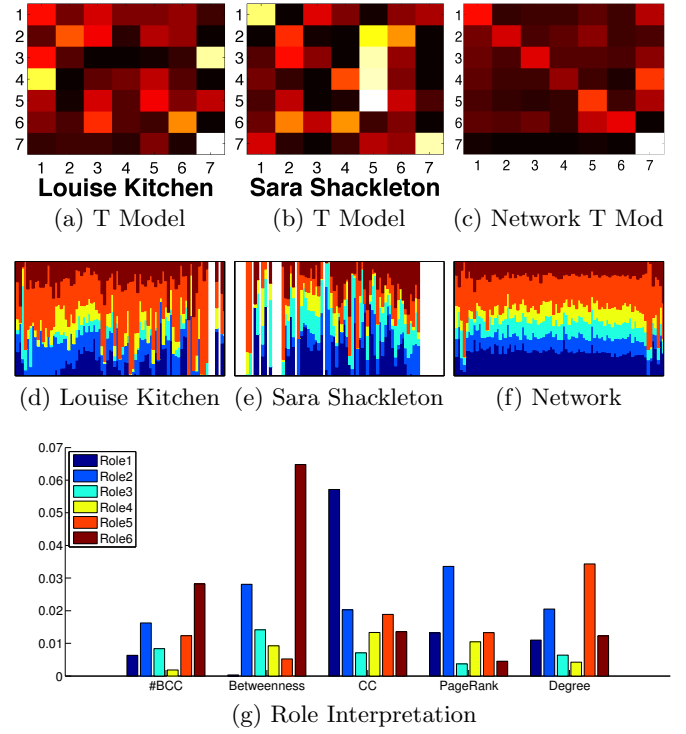
Instead of providing subjective or anecdotal evidence for what the roles represent, we interpret the roles of the DBMM with respect to well-known node measurements such as degree, pagerank, clustering coefficient, betweenness, and bi-connected components. We extend the analytical tools from [16] for use in interpreting the role dynamics. The first technique interprets the roles using the dynamic node-role memberships  $G_t$  and a node measure matrix  $M_t \in \mathbb{R}^{n \times m}$  to compute a non-negative matrix  $E_t$  such that  $G_t E_t \approx M_t$ . The node measurements used are betweenness, biconnected components, pagerank, clustering coefficient, and degree. The matrix  $E_t$  represents the contributions of the node measures to the roles at time  $t$ . We report average contributions over time. Figure 3(c) shows this quantitative interpretation of roles for the IP network. Intuitively, the first role represents degree, while the third and fourth roles represent mainly nodes with high pagerank whereas role five represents nodes with high clustering coefficient and role six represents betweenness. The other roles represent more specialized structural motifs as they are represented by a combination of structural characteristics.

#### 4.2.2 Anomalous Temporal Behavior

We further demonstrate the use of DBMMs for detecting anomalies in time-evolving networks. In particular, we formulate this problem with respect to identifying nodes that have unusual structural transition patterns. For instance, a node might transition from being a hub (i.e., a node with many people linking to it) to a node with high clustering coefficient.

**Node Anomaly.** While there are many ways to define an anomaly detection technique with respect to the DBMM model, we propose an intuitive algorithm shown in Alg. 1 that uses a node’s transition model for predicting the network memberships at  $t + 1$ . The anomaly score is the difference between the predicted network mixed-memberships and the ground-truth mixed-memberships. Therefore, the score represents the divergence of that nodes transitions from the entire network. One simple example is shown in Figure 4(a) where we find Louise Kitchen as having unusual behavioral transitions.

**Time-varying Node Anomalies.** For detecting the specific time interval in which a node has unusual behavior we use the previous method with a few subtle distinctions. The global and node models are estimated at each *timestep* (in a sort of streaming fashion) using the stacked representation with a shorter window (for leveraging past training examples). The final result is a ranked list of potential node anomalies for each timestep, shown in Figure 6. The justification for such an approach is that nodes may become anomalous or have unusual behavior only for a specified time interval. In the case of IP-communications, it is unlikely for the behavior of a node/IP address to remain unusual as IP-addresses are released/expires and users are assigned en-



**Figure 4:** The DBMM transition model provides an effective means of automatically discovering and visualizing nodes with anomalous temporal behavior. 4(a) and 4(b) compare the transition models for two of the most anomalous nodes in the Enron email network to the normal network transition model in 4(c). 4(d)-4(f) show the corresponding role memberships over time. 4(g) shows the characteristics of individual roles.

tirely new IP-addresses. These types of dynamic anomalies are shown in Figure 6.

**Anomalous Structural Transitions.** We first interpret the roles and their temporal variation quantitatively as shown in Figure 4(g) and then provide some simple examples of nodes that have unusual behavior transitions. Intuitively, the first role represents nodes with high clustering coefficient, the second role represents mainly nodes with high pagerank, while the third and fourth roles represent some type of combination of these properties indicating a more complex structural motif that is not sufficiently represented by the selected node metrics. However, the fifth role repre-

---

#### Algorithm 1 Anomalous Structural Transitions

---

**Input:**  $\mathbf{G} = \{G_t : t \in \mathcal{T}\}$  (evolving mixed-memberships)  
**Output:** Ranked list of anomalies and scores

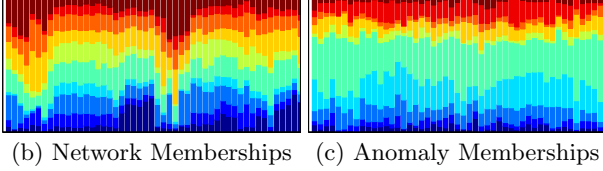
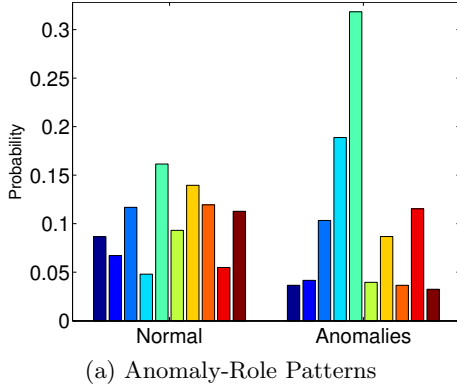
---

```

1: for  $i = 1 \rightarrow |\mathbf{G}|$  do
2:   for  $t = 2 \rightarrow |\mathcal{T}|$  do
3:      $W_{t-1} = G_{t-1}(i)$ 
4:      $W_t = G_t(i)$ 
5:   end for
6:    $T_t(i) \in \mathbb{R}^{r \times r} \leftarrow NMF(W_{t-1}, W_t)$ 
7:    $\hat{G}_{t+1} = T_t(i) \cdot G_t$ 
8:    $\phi(i) = \|\hat{G}_{t+1} - G_{t+1}\|_F$ 
9: end for
```

---





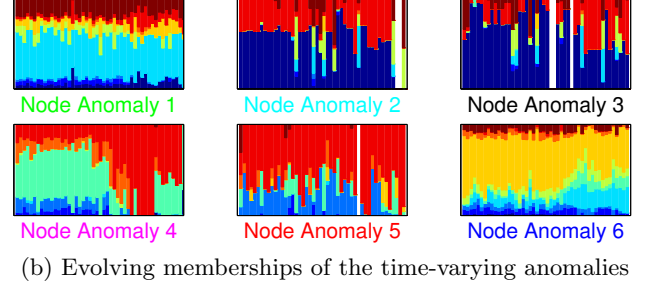
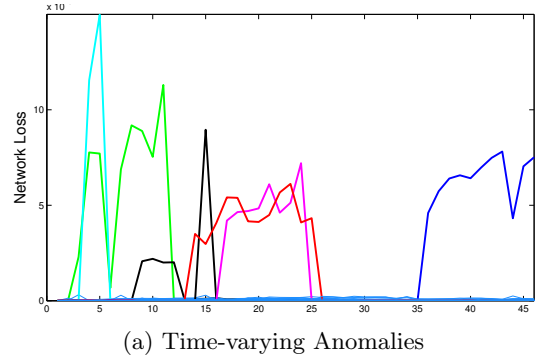
**Figure 5:** The DBMM anomaly detector effectively captures differences in both static and dynamic behavior in an email network. 5(a) shows that normal and anomalous nodes differ in their role distribution (i.e., overall static behavior). 5(b)-5(c) show that normal and anomalous nodes also differ in how their behavior changes over time, with anomalous nodes exhibiting more stable behavior over time than normal.

sents nodes with high degree and the sixth role represents nodes that are articulation points or that have high betweenness. Additionally, by analyzing the neighbors roles dynamically, we find that nodes with high clustering coefficient primarily are neighbors to nodes with high betweenness or high degree (this plot has been removed for brevity).

In Figure 4(e), we find Louise Kitchen, one of the Enron executives who was involved in the Fraud, as having unusual behavioral transitions. Further examination of the network transition model and the average evolution of the behavioral mixed-memberships provide further insights into his abnormal activities. In particular, there are two main role transitions ( $r_3 \rightarrow r_7$  and  $r_4 \rightarrow r_1$ ) in Louise Kitchen’s transition model that are in contradiction with the network transition model shown in Figure 4(c). Furthermore, analyzing the individual changes to the mixed-memberships over time compared with the average behavioral mixed-memberships provides additional insight. For instance, the first two mixed-memberships vectors of Louise Kitchen are mainly red and then begin to deviate significantly with seemingly no underlying correlation or pattern between the role transitions. Moreover, there is not any significantly correlation between Louise’s mixed-memberships and the average mixed-memberships at each timestep.

In addition, we also identify interesting patterns of the nodes with unusual behavioral transitions in Figure 5. In particular, the DBMM anomaly detector effectively captures differences in both static and dynamic behavior in an email network. 5(a) shows that normal and anomalous nodes differ in their role distribution (i.e., overall static behavior). 5(b)-5(c) show that normal and anomalous nodes also differ in how their behavior changes over time, with anomalous nodes exhibiting more stable behavior over time than normal.

Figure 6 also indicates that the DBMM model allows us to



**Figure 6:** The DBMM model allows us to find nodes that are anomalous for only short periods of time and normal otherwise. Such temporally local anomalies are often impossible to find using static graph analysis because brief abnormal periods are drowned out by mostly normal behavior. 6(a) shows examples of short lived anomalies in a computer network. 6(b) shows the corresponding behavior over time for each node in detail.

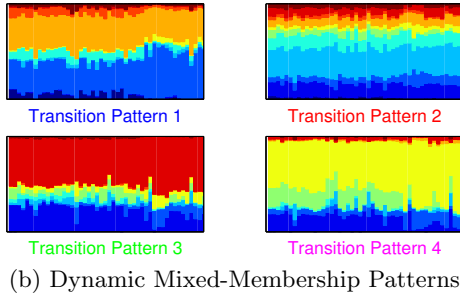
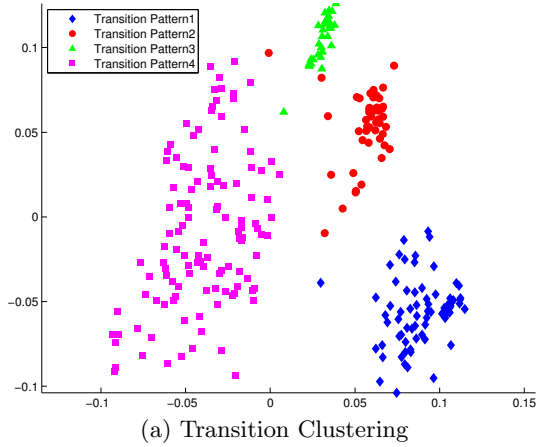
find nodes that are anomalous for only short periods of time and normal otherwise. Such temporally local anomalies are often impossible to find using static graph analysis because brief abnormal periods are drowned out by mostly normal behavior. 6(a) shows examples of short lived anomalies in a computer network. 6(b) shows the corresponding behavior over time for each node in detail.

#### 4.2.3 Clustering Temporal Transitions

Lastly to show the patterns of learned transition matrices, we cluster nodes based on their temporal transition patterns. We find that this clustering reveals the underlying structural patterns of the evolving mixed-memberships. Formally, let  $T^{(i)}$  and  $T^{(j)}$  be the transition matrices of two nodes  $i$  and  $j$ . Then we create an  $r \times r$  vector from each of the node transition models and define a similarity function between these vectors.

First we estimate a single transition model  $T$  for each node using the stacked model. We then compute an  $n \times n$  similarity matrix using Frobenious loss between the transition matrices from the nodes. Next, we apply the classical k-means clustering algorithm to cluster the nodes by their transition matrices. Afterwards, we compute the closest rank- $k$  approximation ( $k = 2$  or  $3$ ) of the similarity matrix. The nodes are plotted using the low-rank approximation and labeled using the previous clustering algorithm. To reveal the structural transition pattern, we then compute the average dynamic mixed-membership for each cluster using only the nodes from that cluster.

This clustering method reveals common structural trends



**Figure 7: The DBMM model provides an intuitive means of clustering nodes that exhibit similar patterns of behavior over time. (a) identifies four distinct clusters of nodes with similar transition patterns. (b) provides a sense of the behavior of each cluster in terms of the average role-membership over time. Again, we see that DBMM captures differences in both overall static behavior (i.e., the specific roles that dominate) and in patterns of how behaviors (i.e., roles) change over time.**

and patterns between nodes. For instance, this technique may group nodes together that share similar transitional patterns such as nodes with stable roles vs. nodes with more dynamic roles or nodes with high activity vs. nodes with low activity. An example is provided in Figure 7. For clarity in the visualization, we randomly selected a small subset of nodes from the 183,389 candidates and identified common transition patterns among them. The first visualization in Figure 7(a) identifies four distinct well-separated clusters of nodes with similar transition models. Figure 7(b) shows the average dynamic behavioral mixed-membership for each cluster. This visualization shows that each cluster represents a unique structural transition pattern between the nodes. The structural patterns can be interpreted using the previous role interpretation from Figure 3(c). This technique can be used for general exploratory analysis such as characterizing the patterns and trends of nodes or eventually used as a means to detect anomalies or nodes that do not fit any transition pattern.

### 4.3 Scalability and Complexity

The temporal behavior model is linear in the number of edges. The complexity can be stated as  $O(|E| \cdot |T|)$  where  $|T|$  is some trivial factor (even in the case where we use minute timesteps for analyzing IP-traces). A more accurate upperbound on the complexity of our model can be

defined in terms of the maximum number of edges at any given timestep. In this case, we can state the complexity as,  $O(|E|_{max(t)} \cdot |T|)$ , where  $|E|_{max(t)}$  denotes the timestep with the maximum number of edges.

Our model is capable of handling realistic networks consisting of billions of nodes and edges. A few examples of these are social networks, communication networks, citation networks, among many others. This is in contrast to similar dynamic mixed-membership models that have been recently proposed such as the dMMSB [30, 10]. These models are quadratic in the number of nodes and thus unable to handle large networks. These models have been typically investigated on trivial sized networks of 18 nodes up to 1,000 nodes. Therefore, these models are unable to scale to the realistic networks with the number of nodes and edges in the billions.

Moreover, the dMMSB can handle 1,000 nodes in a day [30] (See page 30), while our model handles  $\approx 8,000$  nodes in 506.61 seconds (or 8 minutes and 26 seconds) shown in Table 2. We provide performance results for other larger datasets of up to 183,389 nodes and 1,631,824 edges. In all cases, even for these large networks with over a million edges, our model takes less than a day to compute and the performance results show the linearity of our model in the number of edges. For recording the performance results, we applied our model to large datasets using a commodity machine Intel Core i7 @2.7Ghz with 8Gb of memory.

Furthermore, the proposed dynamic behavioral mixed-membership model is also trivially parallelizable as features, roles, and transition models can be learned at each timestep independent of one another. This parallelization makes our model even more attractive and applicable for real-time analysis of the trends and patterns of communication and social networks.

## 5. RELATED WORK

There has been an abundance of work in analyzing dynamic networks. However, the majority of this work focuses on dynamic patterns [11, 18, 17, 23, 26], temporal link prediction [9], anomaly detection [2], dynamic communities [19, 27, 12], and many others [31, 13].

In contrast, we propose a scalable dynamic mixed-membership model that captures the node behaviors over time and consequently learns a predictive model for how these behaviors evolve over the time. Perhaps the most related work is that of [10] where they develop the dMMSB model to discover roles in the graph and how these memberships change over time. However, this type of mixed-membership model assumes a specific parametric form, which is not scalable (1,000 nodes takes a day to model), and where the groups are defined through linkage to specific nodes (in particu-

**Table 2: Performance Analysis of the Dynamic Behavioral Mixed-Membership Model. The dMMSB takes a day to handle 1,000 nodes [30], while our model takes only 8.44 minutes for 8,000 nodes.**

Dataset	N	E	Performance
ENRON	151	50,572	117.51 seconds
TWITTER (COPEN)	8,581	27,889	506.61 seconds
FACEBOOK	46,952	183,831	1,468.65 seconds
INTERNET AS	37,632	505,772	1,922.85 seconds
NETWORK-TRACE	183,389	1,631,824	16,138.71 seconds

lar types of groups) rather than more general node behavior or structural properties [30]. This is in contrast to our proposed model, which is based on our intuitive behavioral representation and can be interpreted quantitatively. In addition, our model is not tied to any single notion of behavior and thus is flexible in the roles discovered and generalizable. Moreover, not only do we evaluate our model on detecting unusual behavior, identifying explainable patterns and trends, and for clustering nodes with respect to their transition patterns, but we apply our model on large real-world networks to demonstrate its scalability. To the best of our knowledge, our proposed model is the first scalable dynamic mixed-membership model capable of identifying explainable patterns and trends on large networks.

## 6. CONCLUSIONS

We proposed a dynamic behavioral mixed-membership model for large networks and used it for identifying interesting and explainable patterns and trends. Moreover, we demonstrated its scalability on a variety of real-world temporal networks and provided striking performance results. The experiments have shown the scalability, flexibility, and effectiveness of our model for identifying interesting patterns, detecting unusual structural transitions, and predicting the future structural changes of the network and individual nodes.

## 7. ACKNOWLEDGEMENTS

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 (LLNL-TR-514271).

## 8. REFERENCES

- [1] Enron email dataset. In <http://www.cs.cmu.edu/enron/>.
- [2] J. Abello, T. Eliassi-Rad, and N. Devanur. Detecting novel discrepancies in communication networks. In *ICDM*, 2010.
- [3] S. Asur, S. Parthasarathy, and D. Ucar. An event-based framework for characterizing the evolutionary behavior of interaction graphs. In *SIGKDD*, 2007.
- [4] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group formation in large social networks: membership, growth, and evolution. In *KDD*, 2006.
- [5] M. Cha, A. Mislove, and K. P. Gummadi. A Measurement-driven Analysis of Information Propagation in the Flickr Social Network. In *WWW*, 2009.
- [6] D. Chakrabarti, R. Kumar, and A. Tomkins. Evolutionary clustering. In *KDD*, 2006.
- [7] A. Dhamdhere and C. Dovrolis. Ten years in the evolution of the internet ecosystem. In *SIGCOMM*, 2008.
- [8] A. Dhamdhere and C. Dovrolis. The internet is flat: modeling the transition from a transit hierarchy to a peering mesh. In *In CoNEXT*, page 21, 2010.
- [9] D. Dunlavy, T. Kolda, and E. Acar. Temporal link prediction using matrix and tensor factorizations. *TKDD*, 5(2):10, 2011.
- [10] W. Fu, L. Song, and E. Xing. Dynamic mixed membership blockmodel for evolving networks. In *ICML*, pages 329–336. ACM, 2009.
- [11] M. Gotz, J. Leskovec, M. McGlohon, and C. Faloutsos. Modeling blog dynamics. In *ICWSM*, 2009.
- [12] D. Greene, D. Doyle, and P. Cunningham. Tracking the evolution of communities in dynamic social networks. In *ASONAM*, 2010.
- [13] H. Habiba, Y. Yu, T. Berger-Wolf, and J. Saia. Finding spread blockers in dynamic networks. In *ASONAM*, 2008.
- [14] D. Hand and R. Till. A simple generalisation of the area under the roc curve for multiple class classification problems. *Machine Learning*, 45(2):171–186, 2001.
- [15] K. Henderson, B. Gallagher, L. Li, L. Akoglu, T. Eliassi-Rad, H. Tong, and C. Faloutsos. It’s Who You Know: Graph Mining Using Recursive Structural Features. In *SIGKDD*, pages 1–10, 2011.
- [16] K. Henderson, B. Gallagher, L. Li, L. Akoglu, T. Eliassi-Rad, H. Tong, and C. Faloutsos. RolX: Role Extraction and Mining in Large Networks. In *LLNL Tech Report*, 2011.
- [17] J. Leskovec, L. Adamic, and B. Huberman. The dynamics of viral marketing. *TWEB*, 1(1):1–39, 2007.
- [18] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. In *KDD*, 2005.
- [19] Y. Lin, Y. Chi, S. Zhu, H. Sundaram, and B. Tseng. Analyzing communities and their evolutions in dynamic social networks. *TKDD*, 3(2):8, 2009.
- [20] P. Mahadevan, D. Krioukov, M. Fomenkov, X. Dimitropoulos, et al. The Internet AS-level topology: three data sources and one definitive metric. *ACM SIGCOMM CCR*, 36(1):17–26, 2006.
- [21] U. of Oregon RouteViews Project. <http://www.routeviews.org/>.
- [22] R. Pan and J. Saramaki. Path lengths, correlations, and centrality in temporal networks. *Arxiv preprint arXiv:1101.5913*, 2011.
- [23] S. Papadimitriou, J. Sun, and C. Faloutsos. Streaming pattern discovery in multiple time-series. In *VLDB*, 2005.
- [24] R. Rossi and J. Neville. Modeling the evolution of discussion topics and communication to improve relational classification. In *SIGKDD 1st Social Media Analytics Workshop*, 2010.
- [25] U. Sharan and J. Neville. Temporal-relational classifiers for prediction in evolving domains. In *ICDM*, 2008.
- [26] J. Sun, C. Faloutsos, S. Papadimitriou, and P. Yu. Graphscope: parameter-free mining of large time-evolving graphs. In *KDD*, 2007.
- [27] L. Tang, H. Liu, J. Zhang, and Z. Nazeri. Community evolution in dynamic multi-mode networks. In *SIGKDD*, pages 677–685. ACM, 2008.
- [28] B. Viswanath, A. Mislove, M. Cha, and K. Gummadi. On the evolution of user interaction in facebook. In *2nd workshop on Online social networks*, 2009.
- [29] X. Wang, X. Liu, and D. Loguinov. Modeling the Evolution of Degree Correlation in Scale-Free Topology Generators. In *INFOCOM*, 2007.
- [30] E. Xing, W. Fu, and L. Song. A state-space mixed membership blockmodel for dynamic network tomography. *The Annals of Applied Statistics*, 4(2):535–566, 2010.
- [31] J. Yang and J. Leskovec. Patterns of temporal variation in online media. In *WSDM*, 2011.